



Havdahl, K. A., Bishop, S. L., Surén, P., Øyen, A. S., Lord, C., Pickles, A., von Tetzchner, S., Schjølberg, S., Gunnes, N., Hornig, M., Lipkin, W. I., Susser, E., Bresnahan, M., Magnus, P., Stenberg, N., Reichborn-Kjennerud, T., & Stoltenberg, C. (2017). The influence of parental concern on the utility of autism diagnostic instruments. *Autism Research*. <https://doi.org/10.1002/aur.1817>

Peer reviewed version

Link to published version (if available):
[10.1002/aur.1817](https://doi.org/10.1002/aur.1817)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via Wiley at <http://onlinelibrary.wiley.com/doi/10.1002/aur.1817/abstract?>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

The influence of parental concern on the utility of autism diagnostic instruments

Karoline Alexandra Havdahl^{1,2,3}, Somer L. Bishop⁴, Pål Surén², Anne-Siri Øyen^{1,2}, Catherine Lord⁵, Andrew Pickles⁶, Stephen von Tetzchner⁷, Synnve Schjølberg², Nina Gunnes², Mady Hornig^{8,12}, W. Ian Lipkin⁸, Ezra Susser^{8,12}, Michaeline Bresnahan⁸, Per Magnus², Nina Stenberg^{2,9}, Ted Reichborn-Kjennerud^{2,10}, Camilla Stoltenberg^{2,11}

¹Lovisenberg Diaconal Hospital, Oslo, Norway

²Norwegian Institute of Public Health, Oslo, Norway

³MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, UK

⁴University of California San Francisco, San Francisco, USA

⁵Weill Cornell Medical College, New York, USA

⁶King's College London, London, UK

⁷Department of Psychology, University of Oslo, Oslo, Norway

⁸Mailman School of Public Health, Columbia University, New York, USA

⁹Oslo University Hospital, Oslo, Norway

¹⁰Institute of Clinical Medicine, University of Oslo, Oslo, Norway

¹¹Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway

¹²New York State Psychiatric Institute, NY, USA

RUNNING TITLE: Influence of parental concern on ASD instruments

Text (main body): 17 pages (5,779 words). Tables: 3. Figures: 2.

Correspondence to: Karoline Alexandra Havdahl, Department of Children's Health, Norwegian Institute of Public Health, Post box 4404 Nydalen, N-0403 Oslo. Tel: +47 93885426.

E-mail: Alexandra.Havdahl[at]fhi.no.

Grant sponsor: The South-Eastern Norway Regional Health Authority; Grant number: 2012101.

Grant sponsor: National Institutes of Health/ National Institute of Neurological Disorders and Stroke; Grant number: U01 NS047537.

Grant sponsor: K.A.H. is a member of the MRC Integrative Epidemiology Unit, which is supported by the UK Medical Research Council (grant number: MC_UU_12013/1).

Abstract

The parental report-based Autism Diagnostic Interview-Revised (ADI-R) and the clinician observation-based Autism Diagnostic Observation Schedule (ADOS) have been validated primarily in U.S. clinics specialized in autism spectrum disorder (ASD), in which most children are referred by their parents because of ASD concern. This study assessed diagnostic agreement of the ADOS-2 and ADI-R toddler algorithms in a more broadly based sample of 679 toddlers (age 35-47 months) from the Norwegian Mother and Child Cohort. We also examined whether parental concern about ASD influenced instrument performance, comparing toddlers identified based on parental ASD concern (n=48) and parent-reported signs of developmental problems (screening) without a specific concern about ASD (n=400). The ADOS cutoffs showed consistently well-balanced sensitivity and specificity. The ADI-R cutoffs demonstrated good specificity, but reduced sensitivity, missing 43% of toddlers whose parents were not specifically concerned about ASD. The ADI-R and ADOS dimensional scores agreed well with clinical diagnoses ($AUC \geq 0.85$), contributing additively to their prediction. On the ADI-R, different cutoffs were needed according to presence or absence of parental ASD concern, in order to achieve comparable balance of sensitivity and specificity. These results highlight the importance of taking parental concern about ASD into account when interpreting scores from parental report-based instruments such as the ADI-R. While the ADOS cutoffs performed consistently well, the additive contributions of ADI-R and ADOS scores to the prediction of ASD diagnosis underscore the value of combining instruments based on parent accounts and clinician observation in evaluation of ASD.

Key words: ADI-R, ADOS, early diagnosis, screening.

Lay abstract

Diagnostic tools for autism spectrum disorders (ASD) have been developed primarily using groups of children who have been referred to ASD clinics by their parents because of concern about ASD. Therefore, these diagnostic tools may not be appropriate for children brought for assessment of ASD in other ways, such as due to general concerns about development. This study examined the usefulness of ASD diagnostic tools in 679 three-year-olds from Norway, and the influence of parent concern about ASD. The children were selected for assessment of ASD in several ways. Some children were selected because their parents expressed concern that their child might have ASD, while others were selected because their parents reported developmental delays or unusual behaviors even if they were not concerned about ASD. All children received a full diagnostic assessment for ASD, including the Autism Diagnostic Interview-Revised (ADI-R), a parental interview, and the Autism Diagnostic Observation Schedule (ADOS), a clinician observational assessment. The results showed that parental concern about ASD affected the ability of the ADI-R to identify children with ASD, but not the ADOS. In the group whose parents did not have a specific concern about ASD, nearly half of the children with ASD were missed by the ADI-R cutoff. Identification of children with ASD in this group improved when using a lower cutoff. The parental interview and the clinician observational assessment contributed uniquely and additively to the prediction of ASD diagnosis in children both with and without parent concern about ASD.

Background

Early diagnosis of autism spectrum disorder (ASD) is important given that interventions in young children are associated with considerable improvements in symptoms and functioning (Zwaigenbaum, et al., 2015). However, the time lag from first evaluation to ASD diagnosis can be long, often more than a year (Crane, et al., 2016; Wiggins, et al., 2006; Zuckerman, et al., 2015). Therefore, reliable and valid instruments are crucial to aid clinicians in making timely and appropriate diagnoses of ASD in toddlers. Among the most widely used assessment instruments for ASD are the Autism Diagnostic Interview–Revised (ADI-R) (Rutter, et al., 2003b) and the Autism Diagnostic Observation Schedule (ADOS) (Lord, et al., 2000). The ADI-R is a semi-structured caregiver interview, in which a trained interviewer asks questions to elicit detailed descriptions of the child’s social-communication and repetitive behaviors. The ADOS is a standardized, semi-structured observational assessment of social communication and repetitive behaviors and interest, which is administered and scored by a trained examiner. The ADI-R and ADOS have demonstrated good agreement with clinical diagnosis of ASD, especially when used in combination (see systematic review: Charman & Gotham, 2013; Falkmer, et al., 2013). Due to findings of reduced diagnostic validity in certain groups (e.g., toddlers, individuals with very low IQ), the ADI-R and ADOS algorithms have recently been revised to better account for influences of age and language level on ASD symptom ratings (Kim & Lord, 2012b; Lord, et al., 2012). The new ADI-R Toddler and ADOS-2 algorithms have shown improved diagnostic agreement compared to the previous algorithms (Kim & Lord, 2012a).

The ADI-R and ADOS are heavily relied upon for diagnostic evaluations of ASD across a range of clinical and research settings worldwide, yet only a few studies have examined their validity outside of ASD specialty clinics in the United States (U.S.). In a Swedish sample of toddlers, Zander et al. (2015) found that the ADOS performed similarly as in the U.S. validation studies, whereas the ADI-R performed differently (i.e., generally lower scores, resulting in increased specificity and reduced sensitivity). The ADI-R algorithms also showed reduced

diagnostic agreement in another study of toddlers in Europe and Israel (de Bildt, et al., 2015). ADI-R sensitivity was especially low among toddlers with ASD using phrase-speech; nearly half of these toddlers scored in the *little-to-no* concern range. These findings warrant further study given that both the ADI-R and ADOS are widely used in Europe (e.g., in more than 80 percent of ASD diagnostic evaluations of toddlers and preschoolers in Norway; Larsen, 2015). Multiple cultural and contextual factors could contribute to the inconsistent results between U.S. and non-U.S. studies. For example, parent awareness of ASD symptoms and/or inclination to report problematic behavior in their toddlers might be generally lower in European countries compared to in the U.S. (Zander et al., 2015). Additionally, parental report could be influenced by health system differences (e.g., whether an ASD diagnosis is required to be eligible for services).

A notable difference between the U.S. and European studies was sampling from ASD specialized clinics, in which most children are referred by their parents because of concern about ASD, compared to from non-specialized neuropsychiatric clinics or via screening. It is possible that, among toddlers with ASD, those whose parents are concerned about ASD may be more severely impaired than those whose parents have nonspecific concerns. However, this would be expected to also result in differences in the performance of the ADOS (not only the ADI-R); yet the ADOS demonstrated good diagnostic validity in Swedish toddlers referred to a non-specialized neuropsychiatric clinic (Zander et al., 2015) as well as in a U.S. sample of toddlers identified based on community screening for signs of developmental delay (Guthrie, et al., 2013). It is also possible that parents who are concerned about ASD may be more aware of and/or more inclined to report autism-related behaviors than parents who do not suspect ASD, thereby affecting the performance of parent report-based instruments such as the ADI-R. Information about the role of parental concern in influencing the performance of the ADI-R and/or ADOS is needed given widespread use of these instruments outside of ASD clinics (Molloy, et al., 2011). Current best practice guidelines recommend routine screening for ASD in general child psychiatry settings followed by diagnostic assessment if signs of ASD are detected (Volkmar, et al., 2014). Therefore, use of the ADI-R and

ADOS with children initially brought for assessment due to general developmental and behavioral concerns is increasingly common.

To date, no study has compared the psychometric properties of ASD diagnostic instruments among children identified for evaluation of ASD in different ways. This study examined agreement between scores and cutoffs from the ADI-R and ADOS and clinical diagnoses among toddlers recruited from a population-based study that employed multiple methods for identification. First, we aimed to examine diagnostic agreement in this broadly based Norwegian sample and compare these estimates with diagnostic agreement reported in U.S. validation studies carried out in ASD clinics. Second, we examined whether parental concern about ASD influenced the performance of the instruments. This was possible given that a subgroup of toddlers was recruited because their parents were concerned about ASD, whereas other toddlers were recruited because their parents reported behavioral signs associated with ASD (e.g., language delay) *without* a specific concern about ASD. In particular, we were interested in whether parental ASD concern influenced the discriminative utility of the standardized cutoffs and/or the dimensional scores from the ASD assessment instruments.

Methods

The Norwegian Mother and Child Cohort Study

A sample from the Norwegian Mother and Child Cohort Study (MoBa) received diagnostic evaluations for ASD as part of a sub-study, the Autism Birth Cohort (ABC) (Stoltenberg, et al., 2010; Surén, et al., 2014). MoBa is a prospective pregnancy cohort study established by the Norwegian Institute of Public Health in 1999, with nationwide recruitment of mothers in association with routine ultrasound examinations (41% participated) (see Magnus, et al., 2016; 2006; Schreuder & Alsaker, 2014). The current data were derived from quality-assured MoBa data files released in 2014 (v7). The children (n=114,500) were born August 1999 through July 2009. The participants are largely of Norwegian or Scandinavian ethnicity (95%) (Myhre, et al., 2012).

Multiple strategies were used to identify children with possible ASD. One strategy was based on parental concern about ASD, defined as the parent responding yes to the question of whether the child has autistic traits/autism in the MoBa questionnaires (ages 3, 5, and 7 years) and/or by self-referral or professional referral (parental agreement was a prerequisite) to the ABC research clinic for suspected ASD. Another strategy, which did not require parental concern about ASD, was screening for behavioral signs associated with ASD in the 3-year MoBa questionnaire (response rate: 58.6%). The screening consisted of questions about language and social development, as well as the Social Communication Questionnaire (SCQ) (Rutter, et al., 2003a) (see criteria in Appendix 1). The participation rate for assessment based on the 3-year-questionnaire was approximately 50%.

Identification routes also included registered ASD diagnoses in the Norwegian Patient Registry, siblings of referred/screen-positive children, and random selection of controls (flow chart in Appendix 1). The clinical assessments were undertaken in 2005-2012 at Lovisenberg Hospital in Oslo, in collaboration with the Norwegian Institute of Public Health and Columbia University.

Participant flow

Of the 114,500 children in MoBa, 1,033 children were clinically assessed for ASD (not invited n=112,255; declined assessment n=1,212). Since the present study focused on the ADI-R and ADOS in toddlers (age <48 months), children assessed at older ages (n=201) or who did not complete the ADI-R/ADOS (n=153) were excluded (see flow chart in Appendix 2). Hence, the initial sample consisted of 679 toddlers (aged 35-47 months). Children with severe sensory (sight/hearing) and/or motor impairments and/or nonverbal mental age below 10 months (n=14) were excluded, as the instruments have not been validated for children with such impairments (Kim & Lord, 2012a; Lord et al., 2012).

ASD was defined as clinical diagnoses of Autistic Disorder (n=41), Pervasive Developmental Disorder Not Otherwise Specified (n=24), and Asperger's Disorder (n=1)

(Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision, DSM-IV-TR; APA, 2000). These three DSM-IV-TR sub-categories of ASD have been found to be well accounted for by a single ASD category (Frazier, et al., 2012). For comparability with previous studies, Rett's Disorder and Childhood Disintegrative Disorder were not included in the ASD case definition (excluded n=2). Non-ASD diagnoses were assigned to 303 toddlers, primarily language disorders (n=204), intellectual disability (n=38), and attention-deficit/hyperactivity disorder (n=20). The remaining 294 children did not meet criteria for any DSM-IV-TR diagnosis (as shown in the flow chart in Appendix 2, the majority of these were recruited randomly as controls). The sample characteristics are presented in Table 1.

TABLE 1

Measures

The **ADI-R** (Norwegian translation) was administered by trained research assistants who had demonstrated research reliability in using the instrument (Rutter et al., 2003b). The original ADI-R algorithm used in the assessments provides a classification for Autistic Disorder. The ADI-R Toddler algorithms, consisting of ADI-R items that have demonstrated high sensitivity and specificity in toddlers of three separate language levels (nonverbal, single words, phrase-speech) (Kim & Lord, 2012b), were retroactively calculated. Each algorithm provides a cutoff prioritizing sensitivity (clinical cutoff), a cutoff prioritizing specificity (research cutoff), as well as ranges of concern about ASD (the clinical cutoff differentiates “mild-to-moderate” from “little-to-no” concern). The Toddler algorithms were used when analyzing dimensional ADI-R scores. When small subgroup sample sizes necessitated collapsing across the language levels, we used the 10 items applicable to children of all language levels.

The **ADOS** (Norwegian translation) was administered by licensed clinical psychologists who had demonstrated research reliability in using the instrument (Lord et al., 2000).. Revised algorithms from the ADOS-2 (Lord et al., 2012) and calibrated severity scores (CSS, range 1-10;

Gotham, et al., 2009) were calculated retrospectively. The CSS was used when analyzing dimensional ADOS scores.

Age-equivalent scores derived from the Vineland Adaptive Behavior Scales were used to measure expressive **language ability** (Sparrow, 1984). **Language level** was defined by ADI-R item 30 (i.e., nonverbal, single words, phrase-speech or better). **IQ** was measured with standard scores from the Stanford–Binet Intelligence Scales-5th Edition for most participants (SB5; full version:n=401, abbreviated version:n=200) (Roid, 2003). Toddlers with lower developmental levels than required for the SB5 received the Mullen Scales of Early Learning (MSEL;n=56) (Mullen, 1995). To avoid floor effects on the MSEL, the ratio full scale IQ was derived from age-equivalent scores (mental age/chronological age*100) (Bishop, et al., 2011).

Behavior problems not specific to ASD (attention problems/hyperactivity, aggressive behavior, and anxiety; hereafter “behavior problems”) were measured by parental report in the MoBa 3-year-questionnaire (scale derived from the Child Behavior Checklist, Achenbach & Rescorla, 2000) (see validity data in Biele, et al., 2014; Zachrisson, et al., 2013), and by clinician observation in the ADOS section for “Other Abnormal Behaviors”, which is not included in ASD algorithms (Havdahl et al., 2016a).

Behavioral signs without concern about ASD was operationalized as meeting the ABC Study screening criteria for parent-reported behaviors associated with ASD while answering no to the question of whether the child has autistic traits (n=400). The screening criteria (see Appendix 1) required parent concern about *some* aspect of the child’s behavior or development (although not necessarily about ASD). **Parental concern about ASD** was defined as the parent answering yes to the question of whether the child has autistic traits in the MoBa 3-year-questionnaire and/or by self-referral or professional referral (parental agreement was a prerequisite) to the ABC research clinic for suspected ASD (n=48). Most toddlers in the ASD concern group also met the screening criteria based on parent-reported behavioral signs (n=39 of 47, missing information for one child).

Procedure

Informed consent was obtained from caregivers, using forms approved by the Regional Committee for Medical and Health Research Ethics in South-Eastern Norway and the Columbia University Medical Center Institutional Review Board. Participants underwent a comprehensive multi-disciplinary clinical evaluation during the course of one or two days. The assessment team did not have access to information from MoBa-questionnaires or previous evaluations. Separate examiners administered the ADOS and ADI-R, without knowledge of results from the other instrument. Inter-rater reliability on both instruments was continuously monitored. There were multiple other sources of information. A child psychiatrist or other physician administered a parental interview of developmental history and current concerns, a physical exam, and a standardized observation of mother-child play interaction. Parents and daycare staff completed questionnaires covering signs of ASD and other psychiatric disorders, language abilities, and executive functioning (see Surén et al., 2014). Psychiatric symptom assessment included the Preschool Age Psychiatric Assessment (PAPA; Egger, et al., 2006) or the Early Childhood Inventory-4th Edition (ECI-4; Gadow & Sprafkin, 2000). Following the assessment, the team met to review all available information and discuss clinical impressions. In accordance with the current gold standard for diagnosing ASD (Falkmer et al., 2013), licensed and experienced clinicians used clinical judgment informed by the full multi-disciplinary evaluation to assign a consensus best-estimate diagnosis (DSM-IV-TR criteria).

Data analysis

All analyses were performed in SPSS 22.0 (IBM Corp., USA) or STATA 13 (StataCorp LP, USA). Significance level was set at 0.05. Group differences were examined with two-samples *t*-tests and chi-square tests or Fisher's exact tests. Effect sizes are reported as Cohen's *d* (*d*; small: 0.20-0.49, medium: 0.50-0.79, large: ≥ 0.80), or Cramer's *V* (*V*; small: 0.10-0.29, moderate: 0.30-0.49, large: ≥ 0.50).

Agreement between instrument scores and clinical diagnoses was estimated using area under the curve (AUC) from receiver operating characteristic (ROC) analyses, a well-established method for assessing the overall discriminative performance of a dimensionally scored instrument compared against a dichotomously defined reference standard (e.g., clinical diagnosis) (Janes et al., 2009). The ROC curve is a plot of the true positive rate (the proportion of children with ASD diagnosis correctly classified as ASD) against the false positive rate (the proportion of children without ASD diagnosis misclassified by the instrument as ASD), across the complete range of possible cutoffs. The Stata procedure *roccomp* was used to compare AUCs. Logistic regression was used to examine whether scores from the ADOS and ADI-R contributed independently to prediction of ASD diagnoses (odds ratios [ORs] are reported).

Agreement between ADI-R/ADOS classifications and clinical diagnoses was examined by calculating sensitivity (the proportion of children with ASD diagnoses classified as ASD) and specificity (the proportion of children without ASD diagnoses classified as non-ASD). Likelihood ratios (LR) are also reported. *LR+* is the ratio of the probability of scoring above the cutoff in children with ASD to the probability in children without ASD ($\text{sensitivity}/1\text{-specificity}$), and estimates >1 indicate increased probability of ASD (small:2-4, moderate:5-10, large: >10). *LR-* is the ratio of the probability of scoring below the cutoff in children with ASD to the probability in children without ASD ($1\text{-sensitivity}/\text{specificity}$), and estimates <1 indicate reduced probability of ASD (small:0.5-0.3, moderate:0.2-0.1, large: <0.1). For comparability with the U.S. validation study by Kim and Lord (2012a), sensitivity and specificity is reported for the ADOS-2 ASD cutoff and the ADI-R Toddler clinical cutoffs, and by language level as defined by ADI-R item 30 (i.e., no words, single words, phrase-speech or higher).

Prior to assessing the influence of parental ASD concern on instrument performance, we examined how parental ASD concern was associated with other relevant child and parent characteristics (i.e., age, language and cognitive abilities, behavior problems, ASD diagnosis, maternal education). To examine the influence on ADI-R and ADOS performance, we employed

ROC regression methods (Janes & Pepe, 2008; Janes et al., 2009) (Stata procedure *rocreg* with linear covariate adjustment and 1,000 bootstrap resamples). This approach allowed assessment of the influence of parental ASD concern on 1) threshold (cutoff) performance, and 2) overall scale performance (the ROC curve). In the first step, parental ASD concern was entered as a predictor of ADI-R/ADOS scores among children *without* ASD (linear regression coefficients are reported). As measures of the effect size of the influence of parental ASD concern on threshold performance, we report the sensitivity and specificity of the cutoffs in toddlers with parental ASD concern compared to those without. In the second step, parental concern about ASD was assessed as a predictor of the capacity of ADI-R/ADOS dimensional scores to differentiate between toddlers with and without ASD diagnosis. Probit regression coefficients and AUC's are reported. The ROC regression analyses were also carried out with adjustment for age, language abilities, nonverbal IQ and behavior problems.

Results

Instrument performance in the total sample

Discriminative performance of the cutoffs

Among children without any clinical diagnosis (all of whom were using phrase-speech), specificity of the ASD cutoff was 96% for the ADOS and 99% for the ADI-R (98% and 99% when restricting to children recruited randomly as controls). To facilitate comparison with previous studies and because the ADI-R and ADOS are intended for differentiation between ASD and other clinical disorders, diagnostic agreement is detailed below and in Table 2 for the comparison of children with ASD versus non-ASD diagnoses. In this subsample, the ADI-R (clinical) cutoff showed good specificity, but sensitivity was modest. Children meeting the ADI-R cutoff were 4-10 times more likely to be diagnosed with ASD than with a non-ASD disorder (see estimates of LR in Table 2, which are calculated from the ratio of sensitivity and specificity estimates). Scoring below the ADI-R cutoff reduced the probability of ASD diagnosis somewhat among children with single words or less ($LR < 0.3$), but only modestly among children with phrase-speech ($LR = 0.5$).

The ADOS (ASD cutoff) showed well-balanced sensitivity and specificity (89% and 83%, respectively). Meeting the ADOS cutoff increased the probability of ASD diagnosis by a factor of 3-4 in children with single words or less, and by 7 in children with phrase-speech. Conversely, not meeting the ADOS cutoff was associated with greatly reduced probability of ASD diagnosis across language levels ($LR \leq 0.2$).

Requiring toddlers to meet both the ADOS (ASD) and ADI-R (clinical) cutoffs resulted in excellent specificity (95%-100%), and greatly increased the probability of ASD diagnosis across language levels ($LR \geq 14$). For children with single words or less, *not* meeting the combined criterion was associated with moderately reduced probability of ASD diagnosis ($LR \leq 0.3$). Given that sensitivity was low among children with phrase-speech (49%), *not* meeting the combined criterion was associated with little reduction in probability of ASD diagnosis in this subgroup ($LR = 0.5$). The least restrictive criterion, requiring children to meet cutoff on either ADOS or ADI-R, resulted in very high sensitivities (95%-100%). Not meeting this criterion was associated with greatly reduced probability of ASD diagnosis across language levels ($LR < 0.1$). Among children with single words or less, the tradeoff was low specificity (57%-64%), whereas among children with phrase-speech, specificity was good (83%).

TABLE 2

Characteristics of the misclassified children are shown in Appendix 4. Among children ultimately diagnosed with ASD, those missed by the ADI-R (clinical) cutoff (false negatives) had higher intellectual ($d=0.63$, $p=0.02$) and language abilities ($d=0.76$, $p<0.01$). Among children diagnosed with disorders other than ASD, those meeting the ADI-R cutoff (false positives) had lower intellectual ($d=0.75$, $p<0.001$) and language abilities ($d=0.68$, $p<0.01$) and more parent-reported behavior problems ($d=0.77$, $p<0.001$). Although no statistically significant differences were found between children with ASD who received ADOS classifications of non-ASD compared to ASD, the subgroup of false negatives was small ($n=7$). Among children with diagnoses other than ASD, those meeting the ADOS cutoff (false positives) had lower intellectual ($d=0.62$,

$p<0.001$) and language abilities ($d=0.37$, $p=0.02$) and more clinician-observed behavior problems during the ADOS administration ($d=0.79$, $p<0.001$).

Discriminative performance of the dimensional instrument scores

The Pearson's correlation between scores on the ADI-R and ADOS was 0.63 (using the full ADI-R Toddler algorithm: phrase-speech $r=0.56$, single words $r=0.52$, nonverbal $r=0.77$). Dimensional scores from both instruments differentiated well between children with and without ASD. AUC was 0.93 for the ADI-R (95% confidence interval [CI]=0.91-0.96) (ASD versus other disorder: AUC=0.90) and 0.95 for the ADOS (95% CI=0.92-0.97) (ASD versus other disorder: AUC=0.92). There was no statistically significant difference between ADOS and ADI-R AUC scores ($\chi^2=0.47$, $p=0.49$). Scores from the ADI-R and ADOS contributed independently to the prediction of ASD diagnoses in logistic regression (ADOS: OR=1.96, $p<0.001$, ADI-R: OR=1.42, $p<0.001$, $\chi^2(2, N=663)=254.44$, $p<0.001$), with independent contributions from the social communication ($p<0.001$) and repetitive behavior domains ($p<0.04$) of both instruments.

The role of parental ASD concern

As described in the methods, the study groups for these analyses were: 1) parental ASD concern (referral to the ABC research clinic for suspected ASD and/or parental endorsement of autistic traits during screening; $n=48$), and 2) parent-reported behavioral signs of ASD *without* endorsement of autistic traits ($n=400$) (see Flow chart in Appendix). Parental ASD concern was associated with diagnosis of ASD (54% versus 9% of toddlers were diagnosed with ASD in the group with and without parental ASD concern, respectively). Within diagnostic group (ASD and non-ASD), toddlers with and without parental ASD concern were largely comparable with regard to age, IQ, and language abilities (see Table 3). Among toddlers ultimately diagnosed with ASD, those whose parents had concern about ASD had more parent-reported behavior problems ($d=0.56$).

TABLE 3

Influence on cutoff performance

The first step of the ROC regression showed that among toddlers without ASD, parental ASD concern was significantly associated with higher ADI-R scores ($B=2.82$ [95% CI=1.79-3.85], $p<0.001$), but not with ADOS scores ($B=0.25$ [95% CI=-0.45-0.94], $p=0.49$). The association with elevated ADI-R scores remained ($B=2.37$, $p<0.001$) after adjusting for IQ, language abilities, and parent-reported behavior problems (all p -values < 0.01).

The substantial effect size of the influence of parental concern about ASD on ADI-R cutoff performance is shown by the differences in sensitivity and specificity of the ADI-R cutoffs (Figure 1). Among toddlers with parental ASD concern, the ADI-R clinical cutoff had 85% sensitivity (95% CI=65%-96%) and 68% specificity (95% CI=45%-86%). These estimates are largely consistent with those reported in the algorithm development study (Kim & Lord, 2012b: sensitivity 80%-94%, specificity 70%-81%). In contrast, the ADI-R classifications showed considerably lower sensitivity and higher specificity among toddlers without parental concern about ASD (Figure 1). A sizable proportion of the toddlers ultimately diagnosed with ASD scored below the ADI-R clinical cutoff (43%). Sensitivity and specificity of the ADOS cutoffs were similar in toddlers with and without parental concern about ASD (Figure 1).

FIGURE 1

Influence on the discriminative performance of the dimensional instrument scores

Parental ASD concern was associated not only with ASD diagnosis, but also with increased ADI-R scores independent of ASD diagnosis (i.e., among the toddlers without ASD). Hence, the presence or absence of parental ASD concern could affect estimates of the agreement between ADI-R scores and ASD diagnosis (parental concern about ASD may in itself drive a sizable portion of the agreement between ADI-R scores and ASD diagnosis). After adjusting for the influence of parental ASD concern, agreement between ADI-R scores and clinical diagnoses was still good, resulting in $AUC=0.85$ (95% CI=0.80-0.91), and parental ASD concern had no statistically

significant effect on the ROC curve (i.e., the ability of ADI-R scores to distinguish between toddlers with and without ASD) (coefficient=-0.28 [95% CI=-1.14-0.57], $p=0.51$). Furthermore, ADI-R scores contributed independently to the prediction of ASD diagnoses also in the subgroup of toddlers without parental concern about ASD (ADI-R: OR=1.37, $p<0.001$, ADOS: OR=1.99, $p<0.001$, $\chi^2(2, N=400)=130.08$, $p<0.001$). Still, as shown by plotting the balance between sensitivity and specificity across all possible ADI-R cutoffs (Figure 2), different cutoffs were required for optimal balance in toddlers with and without parental concern about ASD (shown in blue and red, respectively).

Since ADOS scores were not significantly influenced by parental ASD concern, there was no need to adjust for this when estimating the ROC curve (AUC=0.93 [95% CI=0.90-0.96]). There was also no statistically significant influence of parental ASD concern on the ability of ADOS scores to distinguish between toddlers with and without ASD (coefficient=0.18[95% CI=-1.30-0.94], $p=0.76$).

FIGURE 2

Sensitivity analyses

In contrast to most previous studies, a few of the children who received clinical diagnoses had been recruited based on random selection ($n=1$ ASD, $n=29$ other non-ASD disorder). The results were very similar when excluding randomly selected children (Appendix 3). Given that the rate of participant exclusion was higher in the group with parental ASD concern compared to the group without, we also carried out the analyses with no exclusions and found essentially unchanged results.

Results were similar when using the unrevised instrument algorithms. The ADI-R Autism criteria performed similarly as the ADI-R Toddler Research cutoff, and the previously proposed relaxed ADI-R ASD criteria (Risi, et al., 2006) performed similarly as the ADI-R Toddler Clinical cutoff. Due to the small sample sizes in subgroups stratified by language level, diagnosis, and parental ASD concern, the analyses of parental ASD concern were performed with collapsed

language levels. The ADOS CSS already takes language level into account, and for the ADI-R we used only the algorithm items applicable to all children across language levels. The results were similar when using a mean item score from the full language-specific algorithms.

Discussion

ASD diagnostic instruments have been validated primarily in children referred for suspected ASD to specialized clinics in the U.S., resulting in standardized thresholds that may not be appropriate for children identified in other ways (e.g., through general population screening). Accordingly, the first aim of this study was to examine diagnostic agreement in a broadly based Norwegian sample. Compared with diagnostic agreement reported in validation studies carried out in ASD clinics in the U.S, we found similar diagnostic agreement for the ADOS cutoffs (85-100% sensitivity and 71-87% specificity). The ADI-R cutoffs showed reduced sensitivity (57-80%) and increased specificity (79-94%). Toddlers with ASD diagnoses who were missed by the ADOS/ADI-R cutoffs had relatively strong intellectual and language abilities. The reverse was found for toddlers with other diagnoses who were misclassified as false positives by the instruments. False positives also tended to have more behavior problems not specific to ASD, such as hyperactivity, irritability and anxiety. These associations appeared to be relatively informant-specific, with parent-reported behavior problems significantly associated with ADI-R misclassifications, and clinician-observed behavior problems significantly associated with ADOS misclassifications. This pattern of informant-specific associations between ratings of ASD symptoms and behavior problems was also found in a recent study of primarily school-aged children from the U.S. (Havdahl et al., 2016a).

Our second aim was to examine whether parental ASD concern influences the performance of the ASD diagnostic instruments. The low sensitivity of the ADI-R cutoffs, especially in toddlers using phrase-speech, is consistent with studies of toddlers recruited primarily through screening and/or referral for general concerns (rather than self-referred due to suspected ASD) (de Bildt et al., 2015; Zander et al., 2015). Comparing toddlers identified by parental ASD concern versus parent-

reported ASD signs (screening) without concern about ASD, the ADOS cutoffs had consistently high sensitivity and specificity. In contrast, the ADI-R cutoffs showed reduced sensitivity and increased specificity among toddlers whose parents did not have a specific concern about ASD. The influence of parental ASD concern on ADI-R scores appeared to be independent of other factors that typically affect ADI-R scores, such as IQ and language level, and of clinician-observed ASD features (ADOS scores). These findings indicate that in addition to child features of ASD, scores on the parental report-based ADI-R are also affected by parents' concern about ASD per se. Even though trained interviewers rate the ADI-R items after eliciting detailed parent descriptions of the child's actual behaviors, parents who are concerned that their child has ASD are likely to be more aware of behavioral features associated with ASD, give more examples and/or provide clearer descriptions of these features. Accordingly, cutoffs derived from specialized ASD clinic settings may miss many children whose parents do not have a specific concern about ASD.

Although the categorical ADI-R cutoffs performed below expectations in parents who were not concerned about ASD, the dimensional ADI-R scores were useful for detecting toddlers with ASD even when parents were not concerned about ASD ($AUC \geq 0.85$). When taking parental concern about ASD into account (i.e., only comparing children with parental ASD concern and children without parental ASD concern) ADI-R scores had comparable accuracy in detecting toddlers with ASD in the two groups (see Figure 2). This suggests that parents who do not have a specific concern about ASD are able to report as diagnostically useful information about their child's social communication and repetitive behaviors as those who are concerned about ASD, but the scores need to be interpreted in light of the lack of ASD concern. Moreover, although the ADOS cutoffs performed well alone, ADI-R scores contributed independently to the prediction of ASD diagnosis over and above the ADOS scores. Parental report remains an important source of information about past and current ASD features beyond what can be observed during clinic visits. The instruments' additive contributions underscore the value of combining direct observation and parent accounts in diagnostic evaluations of ASD (Kim & Lord, 2012a; Zander et al., 2015).

The findings illustrate that sensitivity and specificity of instrument cutoffs can vary widely depending on the characteristics of the sample. The ADI-R cutoffs, which have been derived from samples of children seen in specialized ASD clinics, missed nearly half of toddlers with ASD whose parents did not have a specific concern about ASD. This demonstrates the importance of considering factors that may influence the performance of instruments used in ASD assessment. While the present study focused on the influence of parent concern about ASD, which may affect the performance of parental report based instruments in particular, direct observation based instruments may be influenced by other factors (e.g., child behavior problems in that context). When cutoffs are relied upon without critical clinical judgment, instrument misclassifications can lead to false reassurance, loss of valuable time for appropriate interventions, and/or inappropriate interventions (Havdahl, et al., 2016b). Interventions for ASD in young children are associated with clinically significant improvements in symptoms, skills and functioning (Zwaigenbaum et al., 2015). Loss of time for ASD interventions could potentially increase the likelihood of more significant impairment and support needs later, as brain maturation and behavioral patterns may become more fixed and less plastic at older ages.

The results have significant implications for instrument development and validation efforts. Studies of samples in ASD specialty clinics have been valuable in forming a primary base for the development, validation, and refinement of diagnostic instruments. While this has positively influenced assessment practices in such clinics, caution must be exercised when applying identical practices to other settings. Parents who are not specifically concerned about ASD cannot necessarily be expected to respond to questions about their child's behavior in the same way as parents who have sought an ASD evaluation for their child. Thus, to achieve an optimal balance of sensitivity and specificity, modified criteria may be required (e.g., alternative cutoffs and/or item combinations). Due to the complexity and ramifications of adjusting algorithms and cutoffs, such changes are not warranted on the basis of a single sample. Replication studies are needed to determine whether and how parental concern about ASD should be taken into account at the level

of instrument design. In the meantime, it behooves clinicians and researchers to consider parental concern about ASD when interpreting scores based on parental report of ASD features. For example, it would be useful to keep in mind that the standard cutoffs on the ADI-R may be too high for toddlers whose parents are not specifically concerned about ASD, and that it is important to integrate parent accounts with information from other sources (e.g., direct observation, daycare staff).

Strengths and limitations

Strengths of the present study included the independent administration of the ADOS and ADI-R by separate clinicians blinded to information from the other instrument, developmental history, and previous evaluations. Nevertheless, as in the majority of studies of ADOS/ADI-R validity, the assessment team was not blinded to information from these instruments in determining clinical diagnoses (Kim & Lord, 2012a; Zander et al., 2015). Currently, the most widely accepted gold standard for ASD diagnosis is expert clinical judgment informed by all available information from a multi-disciplinary evaluation that includes multiple sources of information and standardized instruments (Falkmer et al., 2013; Kim, et al., 2015). Beyond the ADOS and ADI-R, the assessment comprised several other sources of information about ASD symptoms, including observation of play interaction, parent interview, and questionnaires completed by parents and daycare staff. Additionally, the assessment team included a specialist clinician who had conducted parent interview and direct child observation, but who was not involved in the administration or scoring of the ADOS or ADI-R. Finally, even though the availability of information from these instruments could have contributed to overestimation of diagnostic agreement, this would be expected both for children with and without parental ASD concern.

DSM-IV-TR diagnostic criteria were used in this study and while findings suggest that the vast majority of children diagnosed under DSM-IV can be expected to receive the same

classification (ASD/non-ASD) under DSM-5 (Huerta, et al., 2012) it is possible that use of DSM-5 criteria could have affected the results.

Whereas this was a relatively large sample of toddlers, sample sizes were small when stratifying by diagnosis and identification method. Therefore, the results should be interpreted with caution and with consideration of their confidence intervals. In addition, the study was conducted in a single culture and generalization to other countries cannot necessarily be assumed. Selection bias could also potentially limit the generalizability of the results. MoBa is population-based, but comparison with the general Norwegian population indicates underrepresentation of young, single and less educated mothers, as well as lower prevalence of certain exposures (e.g., smoking in pregnancy) (Nilsen, et al., 2013; Statistics Norway, 2016). Self-selection bias may also have been associated with participation in the clinical assessment. Nonetheless, the participation rate for invitations based on the 3-year-questionnaire was relatively high (parental ASD concern: 52.7%, parent-reported behavioral signs but no concern about ASD: 50.4%).

Conclusion

In this broadly based sample of toddlers, the ADOS cutoffs showed consistent and well-balanced sensitivity and specificity. However, the ADI-R cutoffs demonstrated low sensitivity, particularly in identifying toddlers with ASD in the absence of specific parental concern about ASD. Different ADI-R cutoffs were needed according to presence or absence of parental ASD concern in order to achieve comparable balance of sensitivity and specificity. When using these different cutoffs, ADI-R scores differentiated toddlers with and without ASD equally well in the presence and absence of parental ASD concern. These results highlight the importance of taking parental concern about ASD into account when interpreting scores from parental report-based instruments such as the ADI-R. Although the ADOS cutoffs performed consistently well, the additive contributions of ADI-R and ADOS scores to the prediction of ASD diagnosis underscore the value of combining instruments based on parent accounts and clinician observation in evaluation of ASD. Future

studies should examine the influence of parental ASD concern on scores from the ADI-R and other parental report-based instruments in other samples, and the utility of adjusting cutoffs and/or algorithms based on whether parents are concerned specifically about ASD.

Acknowledgements

The authors gratefully acknowledge the contribution of the families that participated in the study.

We thank Kari Kveim Lie, M.D., Britt Kveim Svendsen, M.D., and the Autism Birth Cohort Study clinic and research staff for their invaluable contributions to the data collection.

Disclosures

Somer L. Bishop, Catherine Lord and Andrew Pickles receive royalties for instruments they have co-authored (ADOS, ADI-R, SCQ). The other authors declare no conflicts of interest.

References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed., text rev.)*. Washington, DC: American Psychiatric Association.
- Biele, G., Zeiner, P., & Aase, H. (2014). Convergent and discriminant validity of psychiatric symptoms reported in The Norwegian Mother and Child Cohort Study at age 3 years with independent clinical assessment in the Longitudinal ADHD Cohort Study. *Norsk Epidemiologi*, 24(1-2).
- Bishop, S. L., Guthrie, W., Coffing, M., & Lord, C. (2011). Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *American journal on intellectual and developmental disabilities*, 116(5), 331-343.
- Charman, T., & Gotham, K. (2013). Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders - lessons from research and practice. *Child Adolesc Ment Health*, 18(1), 52-63.
- Crane, L., Chester, J. W., Goddard, L., Henry, L. A., & Hill, E. (2016). Experiences of autism diagnosis: A survey of over 1000 parents in the United Kingdom. *Autism*, 20(2), 153-162.
- de Bildt, A., Sytema, S., Zander, E., Bolte, S., Sturm, H., Yirmiya, N., et al. (2015). Autism Diagnostic Interview-Revised (ADI-R) Algorithms for Toddlers and Young Preschoolers: Application in a Non-US Sample of 1,104 Children. *Journal of Autism and Developmental Disorders*, 45(7), 2076-2091.
- Egger, H. L., Erkanli, A., Keeler, G., Potts, E., Walter, B. K., & Angold, A. (2006). Test-retest reliability of the Preschool Age Psychiatric Assessment (PAPA). *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(5), 538-549.
- Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child and Adolescent Psychiatry*, 22(6), 329-340.
- Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., et al. (2012). Validation of proposed DSM-5 criteria for autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 51(1), 28-40 e23.
- Gadow, K., & Sprafkin, J. (2000). *Early Childhood Inventory-4 screening manual*. Stony Brook, NY: Checkmate Plus Ltd.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693-705.
- Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry*, 54(5), 582-590.
- Havdahl, K. A., Hus Bal, V., Huerta, M., Pickles, A., Oyen, A. S., Stoltenberg, C., et al. (2016a). Multidimensional influences on autism symptom measures: Implications for use in etiological research. *Journal of the American Academy of Child and Adolescent Psychiatry*, 55(12), 1054-1063.e1053.
- Havdahl, K. A., von Tetzchner, S., Huerta, M., Lord, C., & Bishop, S. L. (2016b). Utility of the Child Behavior Checklist as a screener for autism spectrum disorder. *Autism Research*, 9(1), 33-42.
- Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2012). Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of

- pervasive developmental disorders. *The American journal of psychiatry*, 169(10), 1056-1064.
- Janes, H., Longton, G., & Pepe, M. (2009). Accommodating covariates in ROC analysis. *The Stata Journal*, 9(1), 17-39.
- Janes, H., & Pepe, M. S. (2008). Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *American Journal of Epidemiology*, 168(1), 89-97.
- Kim, S. H., & Lord, C. (2012a). Combining information from multiple sources for the diagnosis of autism spectrum disorders for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Child Psychology and Psychiatry*, 53(2), 143-151.
- Kim, S. H., & Lord, C. (2012b). New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorders*, 42(1), 82-93.
- Kim, S. H., Macari, S., Koller, J., & Chawarska, K. (2015). Examining the phenotypic heterogeneity of early autism spectrum disorder: Subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry*, [Epub ahead of print].
- Larsen, K. (2015). The early diagnosis of preschool children with autism spectrum disorder in Norway: A study of diagnostic age and its associated factors. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, 3(2), 136-145.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205-223.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Modules 1-4*. Los Angeles, California: Western Psychological Services.
- Magnus, P., Birke, C., Vejrup, K., Haugan, A., Alsaker, E., Daltveit, A. K., et al. (2016). Cohort profile update: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*, 45(2), 382-388.
- Magnus, P., Irgens, L. M., Haug, K., Nystad, W., Skjaerven, R., Stoltenberg, C., et al. (2006). Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *International Journal of Epidemiology*, 35(5), 1146-1150.
- Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism*, 15(2), 143-162.
- Mullen, E. (1995). *Mullen Scales of Early Learning*. Minneapolis, MN: Pearson Assessments.
- Myhre, M. C., Thoresen, S., Grogard, J. B., & Dyb, G. (2012). Familial factors and child characteristics as predictors of injuries in toddlers: A prospective cohort study. *BMJ Open*, 2(2), e000740.
- Nilsen, R. M., Suren, P., Gunnes, N., Alsaker, E. R., Bresnahan, M., Hirtz, D., et al. (2013). Analysis of self-selection bias in a population-based cohort study of autism spectrum disorders. *Paediatric and Perinatal Epidemiology*, 27(6), 553-563.
- Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(9), 1094-1103.
- Roid, G. H. (2003). *Stanford-Binet intelligence scales (5th ed.)*. Itasca, IL: Riverside.
- Rutter, M., Bailey, A., & Lord, C. (2003a). *Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.
- Rutter, M., Le Couteur, A., & Lord, C. (2003b). *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, California: Western Psychological Services.

- Schreuder, P., & Alsaker, E. (2014). The Norwegian Mother and Child Cohort Study (MoBa)– MoBa recruitment and logistics. *Norsk Epidemiologi*, 24(1-2).
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Services.
- Statistics Norway. (2016). Population's level of education Retrieved November 20, 2016, from <https://http://www.ssb.no/statistikkbanken/selecttable/hovedtabellHjem.asp?KortNavnWeb=utniv&CMSSubjectArea=utdanning&PLanguage=1&checked=true>
- Stoltenberg, C., Schjølberg, S., Bresnahan, M., Hornig, M., Hirtz, D., Dahl, C., et al. (2010). The Autism Birth Cohort: A paradigm for gene-environment-timing research. *Molecular Psychiatry*, 15(7), 676-680.
- Surén, P., Schjølberg, S., Øyen, A.-S., Lie, K. K., Hornig, M., Bresnahan, M., et al. (2014). The Autism Birth Cohort (ABC): A study of autism spectrum disorders in MoBa. *Norsk Epidemiologi*, 24(1-2), 39-50.
- Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., State, M., et al. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53(2), 237-257.
- Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental and Behavioral Pediatrics*, 27(2 Suppl), S79-87.
- Zachrisson, H. D., Dearing, E., Lekhal, R., & Toppelberg, C. O. (2013). Little evidence that time in child care causes externalizing problems during early childhood in Norway. *Child Development*, 84(4), 1152-1170.
- Zander, E., Sturm, H., & Bölte, S. (2015). The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: Diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*, 19(2), 187-199.
- Zuckerman, K. E., Lindly, O. J., & Sinche, B. K. (2015). Parental concerns, provider response, and timeliness of autism spectrum disorder diagnosis. *The Journal of Pediatrics*, 166(6), 1431-1439 e1431.
- Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., et al. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics*, 136 Suppl 1, S60-81.

Table 1: Total sample characteristics by expressive language level and diagnosis.

	Phrase-speech						Single words				Nonverbal			
	ASD diagnoses		Other diagnoses		No diagnoses		ASD diagnoses		Other diagnoses		ASD diagnoses		Other diagnoses	
	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)
Age in months	35	41.2 (2.6)	231	41.5 (2.1)	294	42.1 (2.2)	20	41.1 (3.3)	58	41.0 (2.4)	11	41.4 (3.2)	14	41.8 (2.6)
Male sex [%]	32	[91.4]	167	[72.3]	165	[56.1]	17	[85.0]	48	[82.8]	6	[54.5]	10	[71.4]
IQ	34	81.3 (20.5)	230	92.5 (15.0)	294	103.6 (12.2)	18	64.0 (19.9)	57	72.0 (15.9)	11	37.7 (11.2)	13	52.7 (19.9)
Language age	33	29.5 (7.6)	222	33.2 (9.1)	283	43.2 (8.9)	20	16.9 (3.4)	58	18.8 (4.4)	11	7.5 (4.3)	14	11.9 (4.6)
Behavior problems (item mean)														
Parent-reported	35	0.6 (0.2)	223	0.6 (0.3)	286	0.5 (0.2)	20	0.8 (0.3)	56	0.6 (0.3)	11	0.6 (0.2)	13	0.5 (0.2)
Clinician-observed	35	0.5 (0.4)	231	0.2 (0.3)	294	0.2 (0.2)	20	0.4 (0.4)	58	0.3 (0.4)	11	0.5 (0.5)	14	0.2 (0.4)
ADOS-CSS	35	5.9 (2.0)	231	1.8 (1.6)	294	1.3 (1.0)	20	6.1 (2.4)	58	2.3 (1.7)	11	7.8 (1.5)	14	2.6 (2.3)
ADI-R (full diagnostic algorithm)	35	13.6 (7.0)	231	4.0 (4.1)	294	1.7 (2.7)	20	12.2 (5.3)	58	4.8 (4.8)	11	15.1 (7.0)	14	5.0 (4.3)
ADI-R (common items)	35	7.1 (4.1)	231	1.9 (2.2)	294	1.0 (1.6)	20	8.1 (3.5)	58	3.3 (3.1)	11	11.6 (5.2)	14	4.4 (3.7)
Identification route, n [%]														
Parental concern for ASD	13	[37%]	14	[6%]	2	[1%]	9	[45%]	4	[7%]	4	[36%]	2	[14%]
Parent-reported ASD signs without concern about ASD	21	[60%]	187	[81%]	112	[38%]	11	[55%]	52	[90%]	5	[46%]	12	[86%]
Controls/other	1	[3%]	30	[13%]	180	[61%]	0		2	[3%]	2	[18%]	0	
Maternal education ≤12 years, n [%] ^a	9	[26%]	73	[33%]	79	[28%]	10	[50%]	23	[40%]	3	[38%]	5	[39%]

Note: Language level strata are based on ADI-R item 30 (overall level of language). ASD=autism spectrum disorder, ADOS-CSS=standardized comparison scores from the autism diagnostic observation schedule, ADI-R Toddler (common items)=sum of the 10 ADI-R diagnostic items common across language levels (Attention to voice, Direct gaze, Seeking to share enjoyment, Range of facial expressions, Appropriateness of social response, Interest in children, Response to approaches of children, Hand and finger mannerisms, Other complex mannerisms, Unusual sensory interests).

^a Missing on maternal education for n=54, percentage of non-missing is reported.

TABLE 2: Agreement of the instrument classifications with clinical diagnoses of ASD versus other disorders.

	N ASD		N Other disorders		Sensitivity	Specificity	LR+	LR-
	TP	FN	TN	FP	95% CI	95% CI		
Total								
ADOS	59	7	252	51	89%, 79-96	83%, 79-87	5	0.13
ADI-R	44	22	275	28	67%, 54-78	91%, 87-94	7	0.37
ADI-R & ADOS	39	27	294	9	59%, 46-71	97%, 94-99	20	0.42
ADI-R or ADOS	64	2	233	70	97%, 90-100	77%, 72-82	4	0.04
Phrase-speech (PS)								
ADOS	31	4	200	31	89%, 73-97	87%, 82-91	7	0.13
ADI-R	20	15	216	15	57%, 39-74	94%, 90-96	9	0.46
ADI-R & ADOS	17	18	225	6	49%, 31-66	97%, 94-99	19	0.53
ADI-R or ADOS	34	1	191	40	97%, 85-100	83%, 77-87	6	0.03
Single words (SW)								
ADOS	17	3	42	16	85%, 62-97	72%, 59-83	3	0.21
ADI-R	16	4	46	12	80%, 56-94	79%, 67-89	4	0.25
ADI-R & ADOS	14	6	55	3	70%, 46-88	95%, 86-99	14	0.32
ADI-R or ADOS	19	1	33	25	95%, 75-100	57%, 43-70	2	0.09
Nonverbal (NV)								
ADOS	11	0	10	4	100%, 72-100	71%, 42-92	4	<0.01
ADI-R	8	3	13	1	73%, 39-94	93%, 66-100	10	0.29
ADI-R & ADOS	8	3	14	0	73%, 39-94	100%, 77-100	na	0.27
ADI-R or ADOS	11	0	9	5	100%, 72-100	64%, 35-87	3	<0.01

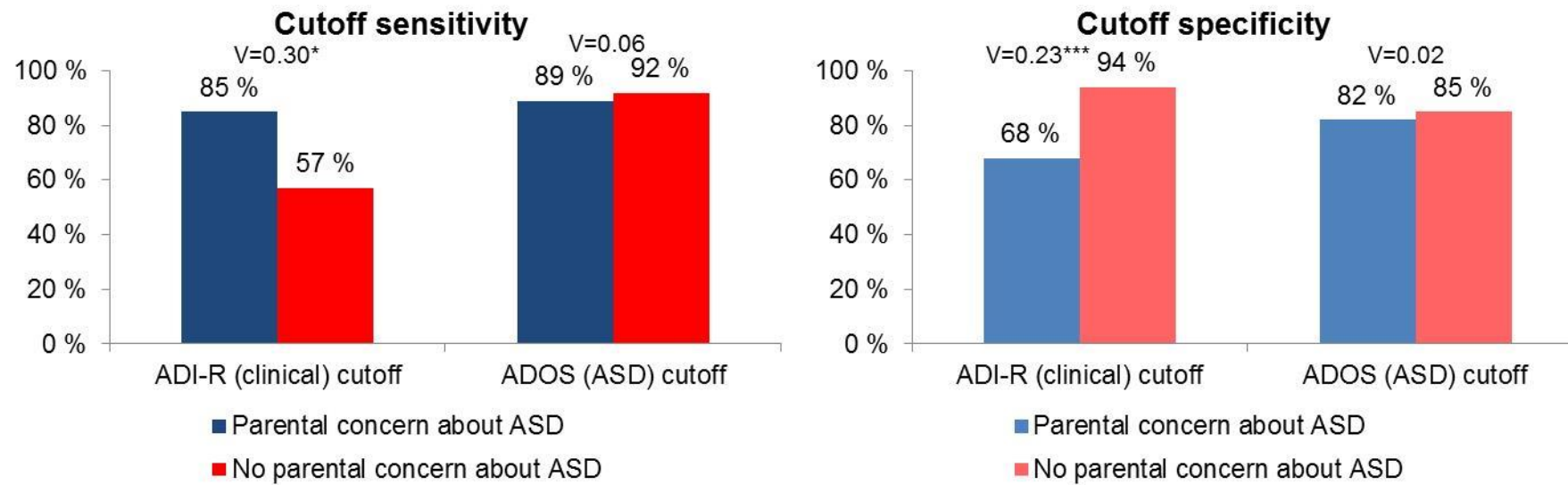
ADI-R=autism diagnostic interview-revised, ADOS=autism diagnostic observation schedule, ASD=autism spectrum disorder, LR+=positive likelihood ratio, LR-=negative likelihood ratio. ADI-R (clinical) cutoff: NV=11, SW=8, PS=13. ADOS-2 (ASD) cutoff: Module 1-no words=11, Module 1-some words=8, Module 2-phrase-speech=7. TP=true positives, TN=true negatives, FP=false positives, FN=false negatives. Language level is defined by ADI-R item 30 (overall level of language).

TABLE 3: Characteristics of toddlers ultimately diagnosed with and without ASD by level of concern about ASD.

	Diagnosed with ASD					Not diagnosed with ASD				
	Parental concern about ASD				p	Parental concern about ASD				p
	Yes		No			Yes		No		
	N	M (SD)	N	M (SD)		N	M (SD)	N	M (SD)	
Demographics										
Age, months	26	41.5 (3.2)	37	41.2 (2.7)	0.64	22	41.4 (3.0)	363	41.6 (2.1)	0.71
Male sex, [%]	23	[88.5]	29	[78.4]	0.30	14	[63.6]	275	[75.8]	0.20
Maternal education ≤12y ^b	10	[41.7]	12	[33.3]	0.51	7	[41.2]	129	[36.3]	0.69
Developmental level										
IQ ^a	25	70.5 (24.7)	35	68.8 (25.6)	0.80	21	84.4 (17.1)	361	90.3 (18.7)	0.17
Language age ^a , months	25	21.6 (10.3)	36	22.3 (10.2)	0.82	19	27.9 (14.1)	351	31.6 (10.5)	0.15
Phrase-speech, [%]	13	[50.0]	21	[56.8]	0.60	16	[72.7]	299	[82.4]	0.26
Behavior problems										
Parent-reported ^a (range:0-2)	26	0.7 (0.3)	37	0.6 (0.2)	0.03	20	0.6 (0.4)	356	0.6 (0.3)	0.35
Clinician-observed (range:0-2)	26	0.4 (0.4)	37	0.5 (0.5)	0.32	22	0.2 (0.3)	363	0.2 (0.3)	0.61
ASD diagnostic instruments										
ADOS CSS (range:1-10)	26	6.2 (2.3)	37	6.3 (2.0)	0.87	22	2.1 (1.7)	363	1.8 (1.6)	0.49
ADI-R (common items) (range:0-20)	26	9.9 (4.6)	37	6.9 (3.6)	<0.01	22	4.8 (3.3)	363	2.0 (2.3)	<0.01

Note: ASD=autism spectrum disorder, ADOS-CSS=standardized comparison scores from the Autism Diagnostic Observation Schedule, ADI-Diag10=sum of the 10 items common across the Autism Diagnostic Interview-Revised Toddler algorithms (Kim & Lord, 2012b). ^aA few cases had missing information (see N). ^bMissing by subgroup from left to right: n=2, n=1, n=5 and n=8. Proportions are calculated based on non-missing n.

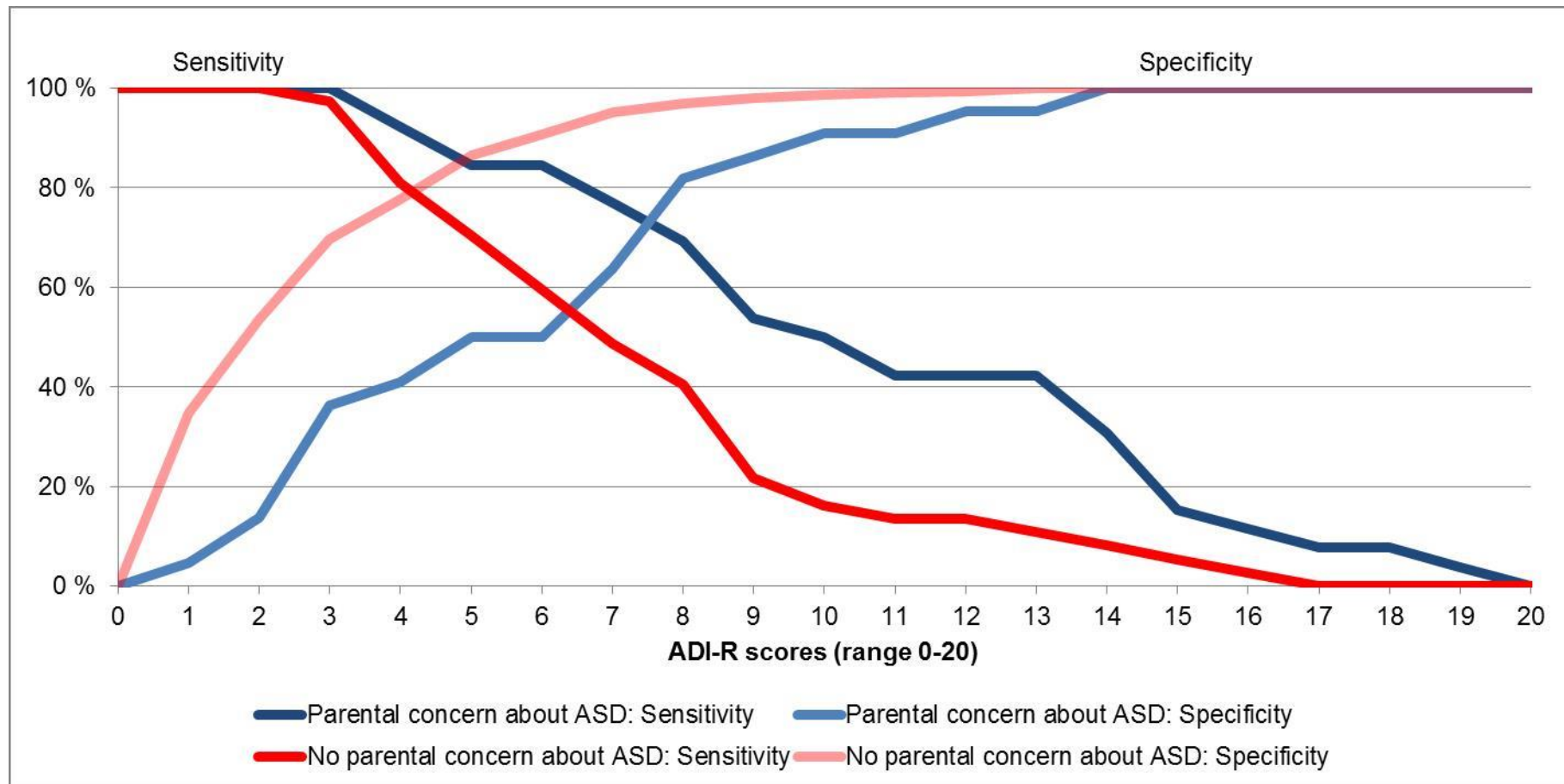
Figure 1: Sensitivity and specificity of ADI-R (clinical) and ADOS (ASD) cutoffs in children with and without parental concern about ASD.



Note: ASD=autism spectrum disorder, ADI-R=Autism Diagnostic Interview-Revised (toddler algorithms), ADOS=Autism Diagnostic Observation Schedule (ADOS-2 algorithms).

V=Cramer's V. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Figure 2: Balance of sensitivity and specificity across ADI-R scores in children with and without parental concern about ASD.



Note: ASD=autism spectrum disorder, ADI-R scores=sum of the 10 items common across the Autism Diagnostic Interview-Revised Toddler algorithms.

Appendix 1: 36-month screening criteria in the Autism Birth Cohort (ABC) Study

Criteria:

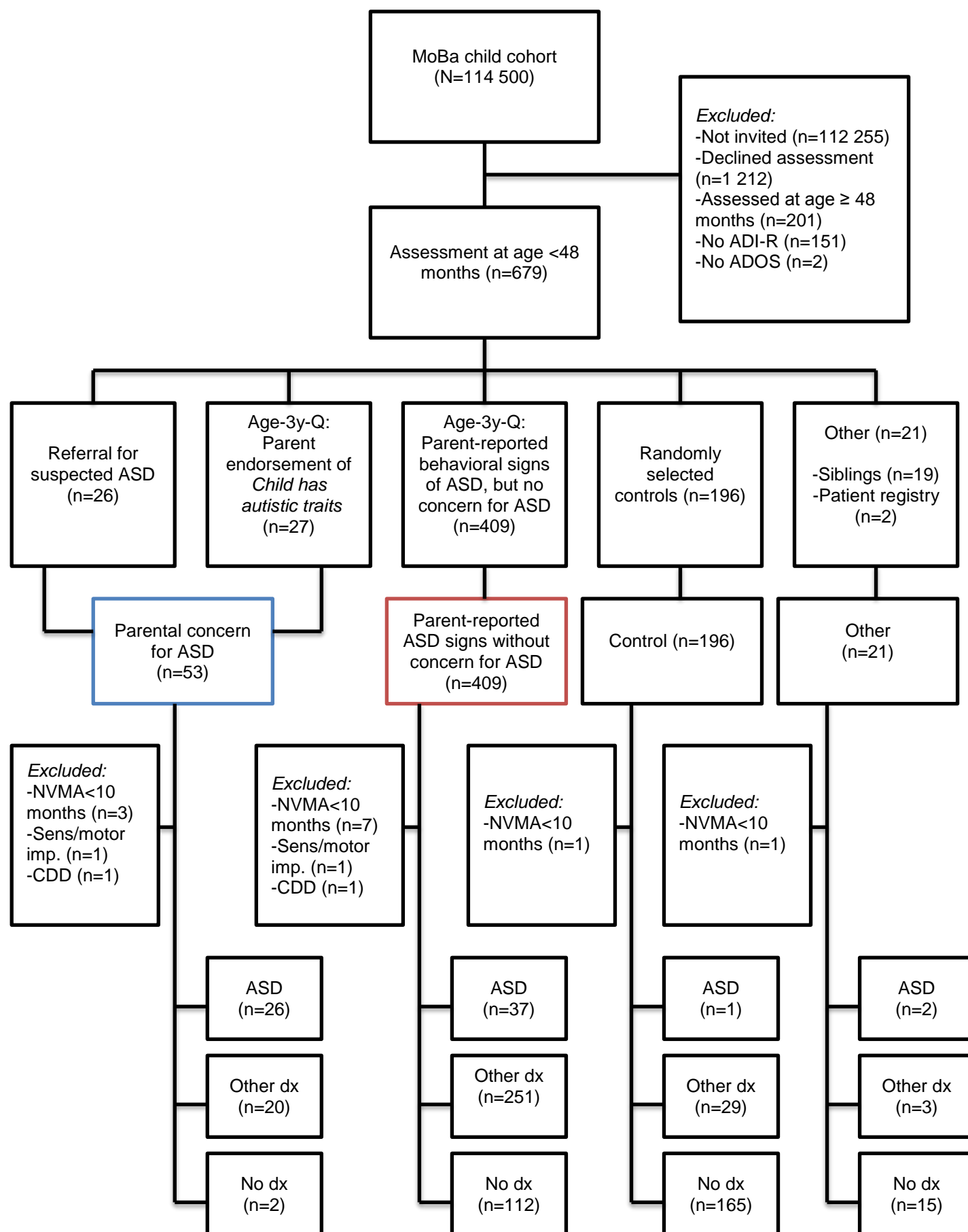
1. Parent reports that the child has autistic traits or has been referred to a specialist for autistic traits
2. SCQ-33 score of ≥ 12
3. Repetitive behavior sub-domain on the SCQ-33 = 9 (full score)
4. Parent reports that the child has been referred to a specialist for language delay
5. Parent reports that the child shows very little interest in playing with other children
6. Parent reports that others (well-baby nurse, teacher, family member) has expressed concern about the child's development

36-month screening algorithm:

- Meeting criterion 1
- Meeting any of criteria 2-5 AND criterion 6

Note: SCQ=Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003). SCQ-33=the 33 of the 40 items which are applicable to both verbal and nonverbal children. Criteria 2 and 3 were based on SCQ scores, while the remaining criteria required the parent to tick off for "yes" when asked about the particular concern.

Appendix 2: Participant flow



ADI-R=autism diagnostic interview-revised, ADOS=autism diagnostic observation schedule, NVMA=nonverbal mental age, Severe sens/motor imp.=severe sensory/motor impairment with profound intellectual disability and autistic traits, CDD=childhood disintegrative disorder. ASD=autism spectrum disorder diagnoses, dx=diagnoses. Colored boxes mark stratification used in analyses of the influence of parental concern for ASD.

Appendix 3: Agreement of the instrument classifications with clinical diagnoses of ASD versus other disorders (excluding children recruited randomly as controls)

	N ASD		N Other disorders		Sensitivity	Specificity	LR+	LR-
	TP	FN	TN	FP	95% CI	95% CI		
Total								
ADOS	59	6	226	48	91%, 81-97	83%, 78-87	5	0.11
ADI-research (res)	30	35	262	12	46%, 34-59	96%, 93-98	11	0.56
ADI-clinical (clin)	44	21	247	27	68%, 55-79	90%, 86-93	7	0.36
ADI-clin & ADOS	39	26	265	9	60%, 47-72	97%, 94-99	18	0.41
ADI-clin or ADOS	64	1	208	66	99%, 92-100	76%, 70-81	4	0.02
Phrase-speech (PS)								
ADOS	31	3	176	28	91%, 76-98	86%, 81-91	7	0.10
ADI-res	11	23	198	6	32%, 17-51	97%, 94-99	11	0.70
ADI-clin	20	14	190	14	59%, 41-75	93%, 89-96	9	0.44
ADI-clin & ADOS	17	17	198	6	50%, 32-68	97%, 94-99	17	0.52
ADI-clin or ADOS	34	0	168	36	100%, 90-100	82%, 76-87	6	<0.01
Single words (SW)								
ADOS	17	3	40	16	85%, 62-97	71%, 58-83	3	0.21
ADI-res	11	9	51	5	55%, 32-77	91%, 80-97	6	0.49
ADI-clin	16	4	44	12	80%, 56-94	79%, 66-88	4	0.25
ADI-clin & ADOS	14	6	53	3	70%, 46-88	95%, 85-99	13	0.32
ADI-clin or ADOS	19	1	31	25	95%, 75-100	55%, 42-69	2	0.09
Nonverbal (NV)								
ADOS	11	0	10	4	100%, 72-100	71%, 42-92	4	<0.01
ADI-res	8	3	13	1	73%, 39-94	93%, 66-100	10	0.29
ADI-clin	8	3	13	1	73%, 39-94	93%, 66-100	10	0.29
ADI-clin & ADOS	8	3	14	0	73%, 39-94	100%, 77-100	na	0.27
ADI-clin or ADOS	11	0	9	5	100%, 72-100	64%, 35-87	3	<0.01

ADI=autism diagnostic interview-revised, ADOS=autism diagnostic observation schedule, ASD=autism spectrum disorder, LR+=positive likelihood ratio, LR-=negative likelihood ratio. ADI-research cutoffs: NV=13, SW=13, PS=16. ADI-clinical cutoffs: NV=11, SW=8, PS=13. ADOS-2 ASD cutoff: Module 1-no words=11, Module 1-some words=8, Module 2-phrase-speech=7. TP=true positives, TN=true negatives, FP=false positives, FN=false negatives.

Appendix 4: Characteristics of misclassified children

	Age		IQ		Language age		Behavior problems not specific to ASD			
							Parent-reported		Clinician-observed	
ADI-R	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)	N	M (SD)
ASD diagnoses – True positive	44	41.1 (3.0)	43	64.0 (25.0)	42	19.2 (9.9)	44	0.7 (0.2)	44	0.5 (0.4)
ASD diagnoses – False negative	22	41.4 (2.6)	20	79.1 (21.5)*	22	26.7 (10.0)**	22	0.6 (0.2)	22	0.5 (0.5)
Other diagnoses – False positive	28	40.9 (2.4)	28	74.3 (19.0)***	27	22.8 (10.3)**	27	0.8 (0.3)***	28	0.4 (0.3) ^a
Other diagnoses – True negative	275	41.5 (2.1)	272	88.2 (18.3)	267	30.0 (10.6)	265	0.5 (0.3)	275	0.2 (0.3)
ADOS										
ASD diagnoses – True positive	59	41.3 (2.9)	56	67.9 (25.3)	57	21.5 (10.7)	59	0.6 (0.2)	59	0.5 (0.4)
ASD diagnoses – False negative	7	40.5 (2.7)	7	75.6 (20.9)	7	24.1 (8.7)	7	0.6 (0.2)	7	0.3 (0.3)
Other diagnoses – False positive	51	41.3 (2.2)	49	77.5 (21.6)***	49	26.0 (11.2)*	48	0.6 (0.2)	51	0.4 (0.4)***
Other diagnoses – True negative	252	41.5 (2.2)	251	88.8 (17.7)	245	30.0 (10.5)	244	0.6 (0.3)	252	0.2 (0.3)

ADI-R= autism diagnostic interview-revised, ADOS = autism diagnostic observation schedule,
 ASD=autism spectrum disorder. Cutoffs: ADI-R Clinical (Kim & Lord, 2012), ADOS ASD (Lord et al., 2012).

*** p<0.001, ** p<0.01, * p<0.05. ^ap=0.050